Cognitive Modelling

David Strohmaier
ALTA Institute, University of Cambridge

Personal Introduction

I am David Strohmaier, a computer scientist and philosopher. Currently, I work as a research associate in the Natural Language and Information Processing (NLIP) group at the University of Cambridge. Concurrently, I am pursuing a PhD in Computer Science at Cambridge. Prevously, I have received a PhD in philosophy from the University of Sheffield.
My research focuses on lexical semantics, but over the years I have worked on many different topics. I am always looking for debate and collaboration. So if you are interested, just contact me!
I am funded by *Cambridge University Press & Assessment*.

Deep Learning and Cognition

A considerable part of my current research turns around the question how much deep learning resembles human cognition. Are there shared universals of cognition that deep neural networks, such as transformers, and human agents both implement? What do the semantic spaces of neural networks capture about human lexical semantics?

Modelling Lexical Semantic Acquisition

Transformer models currently rank as the best performing language models. Their impressive performance in predicting language output raises the question whether they can also be interpreted as cognitive models of the original producers of text, i.e. human language users. After all, the easiest way to predict a word in a textual context might be to recapitulate the cognitive process of the original human producer.
I work on modelling lexical semantic acquisition, i.e. the learning of word meaning, using transformer models. My focus is on second language acquisition of English.

Use Cases	• Identify similarities and differences between transformer models and human language learners. • Model trajectories of language learners as they progress over time. • Support individualised learning applications.

I intend to investigate the resemblance between transformer models and language learner with regard to:

- Real and false cognates
- Semantic errors
- Complex word identification
- Translation equivalence
- Polysemy
- Abstractness

Modelling Contrafactives

Joint work with Simon Wimmer

Factives are attitude verbs that presuppose the truth of the attitude's content. For example:

> *Ahab realised that the whale was close.*

This sentence presupposes that the whale was, in fact, close. A contrafactive would be the inverse, i.e. it presupposes the falsity.

> *Ahab contrad that the whale was close.*

This sentence presupposes that Ahab *wrongly* believed that the whale was close. Interestingly, no human language appears to lexicalise contrafactive attitude verbs.
We propose that one reasons for this is that contrafactives are harder to learn. We test this hypothesis by training transformers models on the semantics of attitude verbs. While small, we found effects supporting our hypothesis.

Publications	• Strohmaier, D., & Wimmer, S. (forthcoming). Contrafactives and Learnability: An Experiment with Propositional Constants. Post-Proceedings of Logic and Engineering of Natural Language Semantics 19. • Wimmer, S., & Strohmaier, D. (2022). Contrafactives and Learnability. In M. Degano, T. Roberts, G. Sbardolini, & M. Schouwstra (Eds.), Proceedings of the 23rd Amsterdam Colloquium (pp. 298–305).

Preference Change

Joint work with Michael Messerli

During my PhD in philosophy I worked on preference change. Together with Michael Messerli, I have a short introduction to that topic forthcoming with Cambridge University Press. The book will address questions such as:

1. What is fundamental preference change?
2. Can we model the dynamics of preference change?
3. How should we choose in light of our changing preferences?

Contact Details

EMail: ds858@cam.ac.uk
Mastodon: https://sigmoid.social/@davidstrohmaier

Check out my website and blog at: *dstrohmaier.com*